

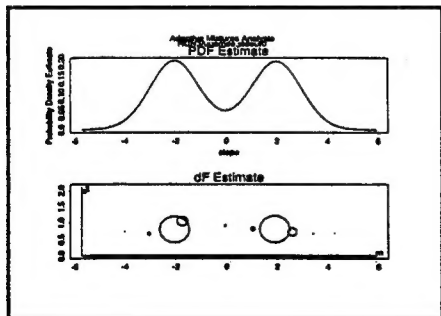
A Method to Determine the Structure of an Unknown Mixture Using the Akaike Information Criterion and the Bootstrap

*Jeffrey L. Solka, Edward J. Wegman, Carey E. Priebe,
Wendy L. Poston and George W. Rogers*

Technical Report No. 105
October, 1994



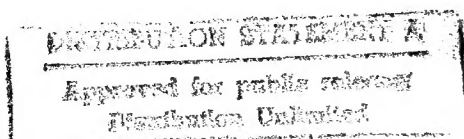
Center for
Computational
Statistics



19941201 028

INFO QUALITY IMPROVED 8

George Mason University
Fairfax, VA 22030



CENTER FOR COMPUTATIONAL STATISTICS
TECHNICAL REPORT SERIES (RECENT REPORTS)

- TR 93. Winston C. Chow, Modeling and Estimation with Fractional Brownian Motion and Fractional Gaussian Noise (Ph.D. Dissertation), February, 1994.
- TR 94. Mark C. Sullivan and Edward J. Wegman, Correlation Estimators Based on Simple Nonlinear Transformations, February, 1994, To appear **IEEE Transactions on Signal Processing**.
- TR 95. Mark C. Sullivan and Edward J. Wegman, Normalized Correlation Estimators Based on Simple Nonlinear Transformations, March, 1994.
- TR 96. Kathleen Perez-Lopez and Arun Sood, Comparison of Subband Features for Automatic Indexing of Scientific Image Databases, March, 1994.
- TR 97. Wendy L. Poston and Jeffrey L. Solka, A Parallel Method to Maximize the Fisher Information Matrix, June, 1994.
- TR 98. Edward J. Wegman and Charles A. Jones, Simulating a Multi-target Acoustic Array on the Intel Paragon, June, 1994.
- TR 99. Barnabas Takacs, Edward J. Wegman and Harry Wechsler, Parallel Simulation of an Active Vision Model, June, 1994.
- TR 100. Edward J. Wegman and Qiang Luo, Visualizing Densities, October, 1994.
- TR 101. Daniel B. Carr, Converting Tables to Plots, October, 1994.
- TR 102. Julia Corbin Fauntleroy and Edward J. Wegman, Parallelizing Locally-Weighted Regression, October, 1994.
- TR 103. Daniel B. Carr, Color Perception, the Importance of Gray and Residuals on a Choropleth Map, October, 1994.
- TR 104. David J. Marchette, Carey E. Priebe, George W. Rogers and Jeffrey L. Solka, Filtered Kernel Density Estimation, October, 1994.
- TR 105. Jeffrey L. Solka, Edward J. Wegman, Carey E. Priebe, Wendy L. Poston and George W. Rogers, A Method to Determine the Structure of an Unknown Mixture Using the Akaike Information Criterion and the Bootstrap, October, 1994.
- TR 106. Wendy L. Poston, Edward J. Wegman, Carey E. Priebe and Jeffrey L. Solka, A Contribution to the Theory of Robust Estimation of Multivariate Location and Shape: EID, October, 1994.
- TR 107. Clifton D. Sutton, Tree Structured Density Estimation, October, 1994.
- TR 108. Charles A. Jones, Simulating a Multi-target Acoustic Array on the Intel Paragon (M.S. Thesis), October, 1994.

A Method to Determine the Structure of an Unknown Mixture Using the Akaike Information Criterion and the Bootstrap

Jeffrey L. Solka¹, Edward J. Wegman², Carey E. Priebe³,
Wendy L. Poston¹, and George W. Rogers¹

¹Dahlgren Division of the Naval Surface Warfare Center
Systems Research and Technology Department
Advanced Computation Technology Group, Code B10
Dahlgren, VA 22448-5000

²Department of Statistics
George Mason University
Fairfax VA 22030-4444

³Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218

Corresponding Author - Jeff Solka
703-663-1982 (W)
703-663-1952 (FAX)
703-663-7226 (SECY)
jsolka@relay.nswc.navy.mil (EMAIL)

ABSTRACT

Given i.i.d. observations $x_1, x_2, x_3, \dots, x_n$ drawn from a mixture of normal terms one is often interested in determining the number of terms in the mixture and their defining parameters. Although the problem of determining the number of terms is intractable under the most general assumptions there is hope of elucidating the mixture structure given appropriate caveats on the underlying mixture. This paper examines a new approach to this problem based on the use of Akaike Information Criterion (AIC) based pruning of data driven mixture models which are obtained from resampled data sets. Results of the application of this procedure to artificially generated and real world data sets are provided.

1. Introduction

Given $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ where each \mathbf{x}_i is d dimensional and i.i.d. according to an unknown density $f_0(\mathbf{x})$ one is often interested in estimating $f_0(\mathbf{x})$. This problem occurs in such areas as exploratory data analysis, classification, and regression. There are a variety of approaches to the multivariate density estimation problem (Scott, 1992).

An often used parametric approach is that of finite mixture models (Everitt and Hand, 1981) in combination with the expectation maximization (EM) method of Dempster, Laird, and Rubin (1977). One difficulty with this tactic is that one needs some idea as to the appropriate number of terms in the mixture model as well as the approximate parameter values. Given this information the EM algorithm is guaranteed to converge to at least a local maxima in the likelihood surface.

Some of the previous nonparametric approaches include histograms (Sturges, 1926), frequency polygons (Scott, 1985a), adaptive histograms (Wegman, 1970), average shifted histograms (Scott, 1985b), and kernel estimators (Silverman, 1986). These approaches are beneficial in that they possess nice asymptotic consistency properties, robustness with regard to nonnormality, and fewer parameters to estimate which implies better estimates in the finite sample regime. They are at a disadvantage as compared to the mixture model approach when it is suspected that the unknown true density is a mixture of a number of terms and one would like to estimate the posteriori probability of underlying term membership for an unlabeled observation.

This type of problem exists in the areas of medical diagnosis and image processing. In medical diagnosis the term membership may play an important role in identification of the

underlying mechanism of disease or the identification of appropriate tissue type (Carmen and Merickel, 1990). In the general problem of image analysis the term membership may pertain to region type.

A recently developed density estimation technique that circumvents some of the problems of the above techniques is the adaptive mixtures procedure of Priebe and Marchette (1993). This procedure is a blend of the finite mixtures and kernel estimator approaches. It is essentially a mixtures type approach that allows for the creation of new terms in a data driven manner. We have successfully applied this technique in combination with fractal-based features to the detection of man-made objects in land (Solka, Priebe, and Rogers, 1992) and aerial (Priebe, Solka, and Rogers, 1993) images, the general problem of texture classification (Solka, Priebe, and Rogers, 1993), and the measurement of breast parenchymal tissue density (Priebe et al., 1994). The adaptive mixtures estimator is asymptotically consistent like the kernel estimator, but it has the added benefit of creating additional terms at a rate which is considerably less than the rate of creation associated with the kernel estimator.

One drawback to the adaptive mixtures estimator is that while there is asymptotic L_1 convergence for the procedure, this convergence is achieved through the creation of an asymptotically infinite number of terms. Thus the procedure will result in an overly complex model given enough data.

In this work we are interested in a model whose complexity more closely matches that of the unknown distribution. The approach is to use the adaptive mixtures procedure as a starting point to generate a mixture model with (potentially many) extra degrees of freedom or parameters and to prune this model to a much smaller mixture model. This pruning

<input checked="checked" type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/>

Date A-1	Special Special
-------------	--------------------

of terms which is based on the use of the Akaike Information Criterion (AIC) is performed to obtain a model that not only matching the underlying distribution in a functional sense but also with regard to model complexity. Subsequent sections will detail how well that our pruning based procedure meets these goals.

The AIC was originally developed as a tool to choose between two statistical estimators of differing complexities (Akaike, 1972). The AIC is written as a function of the likelihood L and number of free parameters M in a model as follows

$$AIC(\hat{f}) = -2\ln(L) + 2M.$$

In Akaike's original paper the AIC was applied to time series analysis, but subsequent work has applied the technique to ISODATA based (Carman and Merickel, 1990) and general clustering (Bozdogan and Sclove, 1984), and finite mixture analysis (Liang Jaszczak, and Coleman, 1992).

We have developed a new approach to finite mixture determination which employs AIC based pruning of AMDE estimates. This approach differs from the work of Liang in two ways. Liang chooses to make an initial guess as to the appropriate complexity of the data's finite mixture model and then adjusts the number of terms in the model up and down by adding and removing a term until no further improvement is possible. Our approach begins with an over determined data driven model that is produced by the AMDE procedure and then uses AIC in combination with the expectation maximization technique to prune superfluous terms from the model. So whereas Liang's approach adds and subtracts terms to the model our approach just removes terms. The second difference is that where Liang's approach is to produce a single best solution to the finite mixtures question our approach produces a distribution of model complexities from which estimates of the appropriate model complexity can be made.

Section 2 develops the methodology for term pruning in the case of finite mixtures models obtained from finite sample application of the adaptive mixtures procedure. Section 3 presents results indicating that we can improve upon an overdetermined mixture model and in some cases determine the true model complexity. Section 4 concludes with a discussion of the relevance of these results.

2. Approach

Our approach combines elements of nonparametric density estimation, parametric density estimation, and information based pruning. The nonparametric AMDE is used as the starting point of our procedure. We begin our discussions with an overview of AMDE.

2.1 Adaptive mixtures density estimation

Given an unknown distribution $f_0(\mathbf{x})$ we seek to model the distribution using $\hat{f}(\mathbf{x})$ defined by

$$\hat{f}(\mathbf{x}; \Psi) = \sum_{i=1}^g \hat{\pi}_i K(\mathbf{x}; \hat{\theta}_i), \quad (1)$$

where K is some fixed density parameterized by $\hat{\theta}_i$, and $\hat{\Psi} = (\hat{\pi}_1, \hat{\theta}_1, \hat{\pi}_2, \hat{\theta}_2, \dots, \hat{\pi}_g, \hat{\theta}_g)$. The $\hat{\pi}_i$'s are referred to as the mixing proportions. (We can assume for much of what follows that K is taken to be the normal distribution, in which case $\hat{\theta}_i$ becomes $\{\hat{\mu}_i, \hat{\Sigma}_i\}$.) In the simplest case the mixture is assumed to have a single term and the parameters that need to be estimated are the mean and covariance of the distribution.

The basic stochastic approach to parameter estimation is to recursively update the estimate $\hat{\Psi}$ of the true parameters Ψ_0 based on the latest estimate $\hat{\Psi}_t$ and the newest data

point \hat{x}_{t+1} . That is,

$$\hat{\Psi}_{t+1} = \hat{\Psi}_t + \Phi_t(\hat{x}_{t+1}; \hat{\Psi}_t) \quad (2)$$

for some update function Φ_t . The specific form of the update equation that we use is the one suggested by Titterton (1984). If we let $I(\Psi)$ be the Fisher information then the version of the recursive update formula we will use is

$$\hat{\Psi}_{t+1} = \hat{\Psi}_t + (nI(\hat{\Psi}_t))^{-1} \left(\frac{\partial}{\partial \hat{\Psi}} \right) \log \hat{f}(\hat{x}_{t+1}; \hat{\Psi}_t) \quad (3)$$

where the derivatives represents the vector of partial derivative with respect to the terms of $\hat{\Psi}$.

In the case of mixtures of multivariate normals we may write the recursive update equations as

$$\hat{\tau}_{n+1}^{(i)} = \frac{\pi_n^{(i)} \hat{f}^{(i)}(\hat{x}_{n+1}; \hat{\theta}_n)}{\sum_{i=1}^g \pi_n^{(i)} \hat{f}^{(i)}(\hat{x}_{n+1}; \hat{\theta}_n)} \quad (4)$$

$$\hat{\pi}_{n+1}^{(i)} = \hat{\pi}_n^{(i)} + \frac{1}{n} \left(\hat{\tau}_{n+1}^{(i)} - \hat{\pi}_n^{(i)} \right) \quad (5)$$

$$\hat{\mu}_{n+1}^{(i)} = \hat{\mu}_n^{(i)} + \frac{\hat{\tau}_{n+1}^{(i)}}{n \hat{\pi}_n^{(i)}} \left(x_{n+1} - \hat{\mu}_n^{(i)} \right), \text{ and} \quad (6)$$

$$\hat{\Sigma}_{n+1}^{(i)} = \hat{\Sigma}_n^{(i)} + \frac{\hat{\tau}_{n+1}^{(i)}}{n \hat{\pi}_n^{(i)}} \left[\left(\hat{x}_{n+1} - \hat{\mu}_n^{(i)} \right) \left(\hat{x}_{n+1} - \hat{\mu}_n^{(i)} \right)^T - \hat{\Sigma}_n^{(i)} \right]. \quad (7)$$

This is where $\hat{\tau}_{n+1}^{(i)}$ is the estimated posteriori probability of \hat{x}_n belonging to the i th term

of the mixture, $\hat{\pi}_{n+1}^{(i)}$ is the estimated mixing coefficient, $\hat{\mu}_{n+1}^{(i)}$ is the d dimensional estimated mean, and $\hat{\Sigma}_{n+1}^{(i)}$ is the dxd estimated covariance matrix of the ith term.

The adaptive mixtures density estimation (AMDE) stochastic approximation approach is to recursively update $\hat{\Psi}$, the estimate of the true parameters Ψ_0 , while at the same time providing the capability to expand the extent of the parameter space $\hat{\Psi}$ if dictated by the underlying complexity of the data. We note that in the AMDE case our parameter space $\hat{\Psi}$ is given by $\hat{\Psi} = (\hat{\pi}_1, \hat{\theta}_1, \hat{\pi}_2, \hat{\theta}_2, \dots)$. The procedure

$$\hat{\Psi}_{t+1} = \hat{\Psi}_t + A \cdot U_t(\hat{x}_{t+1}; \hat{\Psi}_t) + B \cdot C_t(\hat{x}_{t+1}; \hat{\Psi}_t, t), \quad (8)$$

is used to recursively update the density where $A = [1 - P_t(\hat{x}_{t+1}; \hat{\Psi}_t)]$, and $B = P_t(\hat{x}_{t+1}; \hat{\Psi}_t)$. P_t represents a possibly stochastic create decision and takes on values 0 or 1. U_t updates the current parameters using equations (4-7) while C_t adds a new term to the model. As is implicit in the equation, the decision to add a new term is a function of the current data point, our current estimation of the parameters, and time. The time dependence is important in those cases that we wish to anneal the probability of creation as a function of training time. The models produced by the AMDE procedure are good functional estimates, but are typically overdetermined with regard to the number of terms.

2.2 Approaches to AIC based pruning of AMDE generated mixture models

Previous work in the literature has examined the application of the AIC to the determination of the number of terms in a finite mixture (Liang, Jaszczak, and Coleman, 1992). The AIC/n estimates -2 times the expected value of the log likelihood of the estimated model (Akaike, 1972)

$$\frac{AIC}{n} = -2E[\int f_0 \log \tilde{f}]. \quad (9)$$

AIC is defined in terms of likelihood, L, and the number of free parameters in the model,

M, as

$$AIC(\hat{f}) = -2\ln(L) + 2M = -2\ln[\hat{f}(\mathbf{x})] + 2M. \quad (10)$$

One uses the AIC to choose between models of differing complexities by selecting the model with the minimum AIC. This choice is equivalent to maximizing the mean likelihood of the model.

Using this idea as a starting point we have developed a procedure that uses a single or set of adaptive mixtures density estimates and produces a pruned model with a lower complexity. This procedure uses AIC to evaluate the appropriateness of lower complexity models that have been subjected to the iterative EM method. In the iterative EM method the update equation takes the form

$$\hat{\Psi}_{t+1} = \hat{\Psi}_t + \Phi(\vec{X}; \hat{\Psi}_t), \quad (11)$$

where Φ is the update function and \vec{X} is the set of observations. In the case of mixtures of multivariate normals we may write the iterative update equations as

$$\hat{\tau}_{ij} = \frac{\hat{\pi}_i \hat{f}_i(\mathbf{x}_j; \hat{\theta})}{\sum_{i=1}^g \hat{\pi}_i \hat{f}_i(\mathbf{x}_j; \hat{\theta})} \quad (12)$$

$$\hat{\pi}_{ij} = \sum_{j=1}^n \frac{\hat{\tau}_{ij}}{n} \quad (13)$$

$$\hat{\mu}_i = \frac{\sum_{j=1}^n \hat{\tau}_{ij} \mathbf{x}_j}{n \hat{\pi}_i}, \text{ and} \quad (14)$$

$$\hat{\Sigma}_i = \sum_{j=1}^n \frac{\hat{\tau}_{ij} (x_j - \hat{\mu}_i) (x_j - \hat{\mu}_i)^T}{n \hat{\pi}_i} \quad (15)$$

This is where $\hat{\tau}_{ij}$ is the estimated posteriori probability that x_j belongs to term i , $\hat{\pi}_{ij}$ is the estimated mixing coefficient, $\hat{\mu}_i$ is the d dimensional estimated mean vector, and $\hat{\Sigma}_i$ is the $d \times d$ estimated covariance matrix for the i th term.

The steps in our pruning procedure are as follows.

Step 1 - Obtain \hat{f}_g an initial adaptive mixtures approximation to f_0 containing g terms.

Step 2 - Compute the AIC of each of the $g-1$ term models after application of the EM method of equations (12-15) to each of the models.

Step 3 - If $AIC(\hat{f}_{g-1}) < AIC(\hat{f}_g)$ for one of the $g-1$ term models then the pruning process is repeated using this model.

Step 4. Repeat this process of pruning and expectation maximization until no further improvement is possible.

It is important to point out that at each pruning step the remaining terms $\hat{\pi}_i$'s are updated based on their Mahalonobis distance to the pruned term prior to updating with the EM method.

Figure 1 illustrates the pruning process. The log likelihood for the true model, the original ten term model, and the pruned and subsequently expectation maximized models are plotted. In this case the process was able to reduce a ten term model of the mixture $.5N(-2,1) + .5N(2,1)$ to the appropriate two term model. This case will be discussed in Section 3.

[FIGURE 1 SHOULD GO ABOUT HERE]

3. Results

This pruning procedure was tested on data sets drawn from two different bimodal two term distributions, one four mode four term distribution, a standard unimodal normal distribution, and the Buffalo snowfall data (Parzen, 1979), see Figure 2. In each simulated data case 10,000 points were drawn from each distribution. The snowfall data consisted of 63 points.

Twenty-five bootstrap resamples were extracted from each of the data sets using their empirical distributions (Efron and Tibshirani, 1993). These resamples are used in a way that is slightly different from the standard procedure. In standard bootstrapping one uses the resamples to estimate the standard error of a statistic whose standard error is not available in closed form. Our goal in bootstrapping is the production of a distribution on the number of terms in the models after AIC based pruning. This distribution can then be used to estimate the number of terms in the true model.

An up to ten term adaptive mixtures model was created for each of the resampled data sets. Each of these models were then subjected to the AIC based pruning process. This process provides a model complexity distribution based on the data set.

[FIGURE 2 SHOULD GO ABOUT HERE]

In Figures 3a and 3b we present adaptive mixtures solutions for two of the resamplings of the data set drawn from $\alpha(x) = .5 \cdot N(-2, 1) + .5 \cdot N(2, 1)$. We have included dF space plots along with the standard functional representation of the distributions. dF space plots are an effective way to display the terms in a mixture. Each term $\pi_i N(\mu_i, \sigma_i^2)$ is plotted as a circle whose radius is proportional to π_i and whose center is given by (μ_i, σ_i^2) . Where it is hard to

discern the distributional structure from a standard function plot it is quite easy in a dF space plot. We notice that the terms in each of the two solutions are markedly different. This phenomena falls under the adage that there is “more then one way to skin a cat” when producing a functional estimate. We also notice that there are more then the “theoretical” number of terms needed. Each of the models is made up of ten terms. The occurrence of a matching number of terms in each model is the result of our initial constraint on the model complexity. It is important to note that though the terms are different in each solution the location and number of modes is not and that there are terms that are superfluous to the minimal representation of the distribution.

[FIGURE 3 SHOULD GO ABOUT HERE.]

Table 1 illustrates the results of the pruning process. For each of the five distributions we have listed the number of terms in the final pruned models for each of the twenty-five resamples. In case a the procedure converged to the correct solution 11 of 25 times, 7 of 25 times in case b, 17 of 25 times in case c, and 3 of 25 times in case d. The procedure converged to a 3 term solution 4 of the 25 times in the case of the snowfall data. The appropriate solution for the case of the snowfall data will be the subject of later discussions.

[TABLE 1 SHOULD GO ABOUT HERE.]

We may estimate the model complexity through the use of statistical measures on this distribution. For example one could choose the minimal order statistic as the measure of the number of terms in the minimal complexity mixture model that characterizes the data. This choice has the advantage that it represents the lowest complexity model obtainable from the procedure. Alternatively one could use the expected value of the distribution. This choice indicates the average complexity of the mixture models that represent the data set.

Table 2 presents the average L_1 error between the true mixture model and the pruned model for each of the first 4 cases for each of the obtained model complexities. No L_1 results were provide for the snowfall data since the true underlying model is unknown. It is encouraging to note that in each case the minimum average error occurs at the appropriate level of model complexity.

[TABLE 2 SHOULD GO ABOUT HERE.]

The number of modes in the snowfall data has been the topic of continued debate throughout the history of density estimation. Arguments have been made in favor of trimodal (Scott, 1992) and unimodal structure (Scott 1994). In Figure 4a we compare the output of the pruning process for one of the 3 term models to a standard kernel estimator with a bandwidth of 6. The bandwidth of 6 was chosen as an appropriate setting to illustrate the trimodality of the data (Silverman, 1986). The 3 term model has been expectation maximized against the original data in order to make a fair comparison between the two. We note that the two models are very similar in character and specifically with regard to the mode placement. In fact if we drop the bandwidth down to 4 we can obtain a solution that is even closer in character to our mixture model, see Figure 4b.

[FIGURE 4 SHOULD GO ABOUT HERE]

The last thing left to be discussed is the output of the pruning procedure. In Figures 5 a, b, c and 6 a and b we present an expectation maximized adaptive mixture solution along with the output of pruning this solution. We notice that the number of terms in the solution has been reduced from ten to the appropriate number in each case. We also notice that the terms left from the process are in approximately the correct location and have about the right mixing coefficients and variances.

[FIGURES 5 AND 6 HERE]

4. Conclusions

The AMDE procedure provides a data driven method for obtaining a good mixture model density estimate. The convergence properties of the procedure tend to guarantee that the model will be of higher complexity than the true density if the latter is a finite mixture. The exceptions to this occur when the sample size is small enough that too few terms are created by the AMDE. The AMDE thus provides a useful mixture model estimate of an unknown density.

The AIC provides a convenient tool for evaluating appropriate model complexity. It serves as a good “rule of thumb” in choosing between models. Under appropriate conditions it has the capability to help reveal the underlying mixture which generates the data set.

In many cases we have reason to believe that the unknown density is a mixture model but of unknown complexity. Then we are often interested in the structure of the underlying mixture model. It is in this case that AIC based pruning can be used to find not only an “optimal” model but also a distribution of pruned models which provides some knowledge about the true density.

In this paper we have presented a new technique to help determine the unknown structure of a mixture model. This technique uses a set of adaptive mixture solutions that have been subject to AIC based pruning to help determine the minimum complexity mixture model that best characterizes the data. The goal of this technique is the production of a more parsimonious mixture model of an unknown distribution.

This approach embodies the spirit in which the AIC should be used in that one is comparing two maximum likelihood solutions. There is a penalty with regard to computational complexity that occurs in the production of expectation maximized models at each step of the pruning process. However the pruning procedure is highly parallel in nature and we would expect substantial speedups on a MIMD machine.

Future work will focus on extending this technique to multivariate distributions, on a more in-depth analysis of the theoretical underpinnings of the approach, and on parallelization of the procedure. We also plan to pursue better term creation techniques within the AMDE framework. Finally we hope to produce an AMDE like estimator that uses term creations and annihilations.

5. References

- Akaike, H. (1974) "A New Look at the Statistical Model Identification." *IEEE Trans. Auto. Control*, vol.19, pp.716-723.
- Bozdogan, H. and Sclove, S. L. (1984) "Multi-sample cluster analysis using Akaike's information criterion." *Ann. Inst. Statist. Math.* **36**, pp. 163-180.
- Carmen, C. S., and Merickel, M. (1990) "Supervising Isodata with an Information theoretic Stopping Rule." *Pattern Recognition*, Vol. 23, No. 1/2, pp. 185-197.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) "Maximum Likelihood from Incomplete Data via the EM Algorithm." *J. Royal Statist. Soc., Series B* **39** 1-38.
- Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*. Chapman and Hall.

London, UK.

Everitt, B. S. and Hand, D. J. (1981) *Finite Mixture Distributions*. Chapman and Hall. London, UK.

Liang, Z., Jaszczak, R. J., and Coleman, R. E. (1992) "Parameter Estimation of Finite Mixtures Using the EM Algorithm and Information Criteria with Applications to Medical Image Processing." *IEEE Transactions on Nuclear Science*, Vol. 39, No. 4, pp. 1126-1133.

Parzen, E. (1979) "Nonparametric Statistical Data Modeling." *J. Amer. Statist. Assoc.*, 74, pp. 105-131.

Priebe, C. E., Solka, J. L., and Rogers, G. W. (1993) "Discriminant Analysis in Aerial Images Using Fractal Based Features." in *Adaptive and Learning Systems II*, F. A. Sadjadi, Ed., *Proc. SPIE 1962*, pp. 196-208.

Priebe, C. E. and Marchette, D. J. (1993) "Adaptive Mixture Density Estimation." *Pattern Recognition*, Vol. 26, No. 5, pp. 771-785.

Priebe, C. E., Solka, J. L., Lorey, R. A., Rogers, G. W., Poston, W. L., Kallergi, M., Qian, W., Clarke, L. P., Clark, R. A. (1994) "The application of Fractal analysis to Mammo-graphic Tissue Classification." *Cancer Letters*, 77, pp. 183-189.

Scott, D. W. (1985a) "Frequency Polygons." *J. Amer. Statist. Assoc.* 80, pp. 348-354.

Scott, D. W. (1985b) "Average Shifted Histograms: Effective Nonparametric Density Esti-

mation in several Dimensions." *Ann Statist.* 13, pp. 1024-1040.

Scott, D. W. (1992) *Multivariate Density Estimation*. John Wiley and sons, New York, NY.

Scott, D. W. (1994) "A Short Course in Multivariate Density Estimation.", *Interface'94*.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall. New York, NY.

Solka, J.L., Priebe, C.E., and Rogers, G.W. (1992) "An Initial Assessment of Discriminant Surface Complexity for Power Law Features." *Simulation* Volume 58, No. 5, pp 311-318.

Solka, J. L., Priebe, C. E., and Rogers, G. W. (1993) "A Probabilistic Approach to Fractal Based Texture Discrimination." in *Adaptive and Learning Systems II*, F. A. Sadjadi, Ed., *Proc. SPIE* 1962, pp. 209-218.

Solka, J. L., Priebe, C. E., Rogers, G. W., Poston, W. L., Lorey, R. A. (1994) "Maximum Likelihood Density Estimation With Term Creation and Annihilation." presented at and appearing in the proceedings of *Interface '94*.

Sturges, H. A. (1926) "The Choice of a Class Interval." *J. Amer. Statist. Assoc.* 21, pp. 65-66.

D. M. Titterington (1984) "Recursive parameter estimation using incomplete data." *J. Royal Stat. Soc., Ser. B, Vol. 46*, pp 257-267.

Wegman, E. J. (1970) "Maximum Likelihood Estimation of a Unimodal Density Function." *Ann. Statist.* 41 457-471.

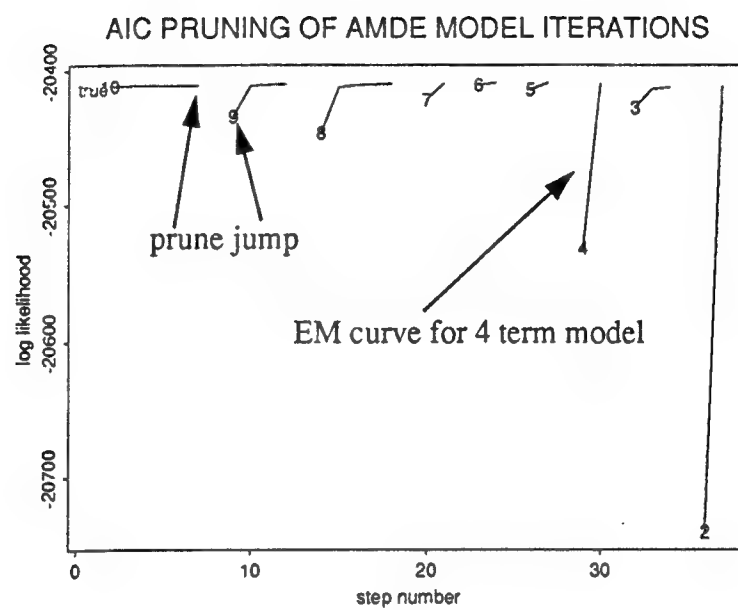
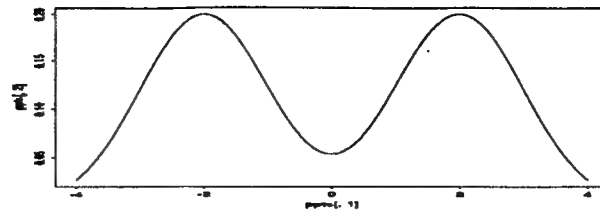
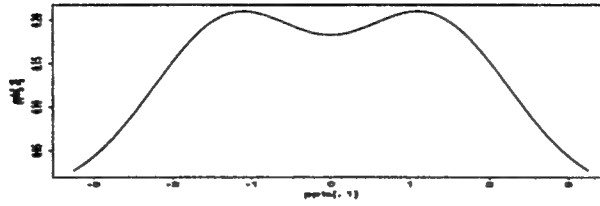


Figure 1

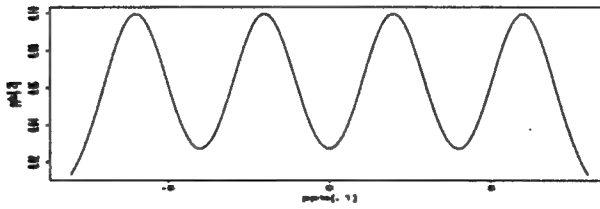
Case a



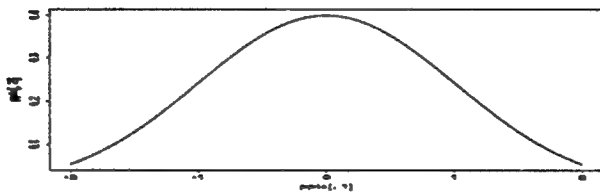
Case b



Case c



Case d



Case e

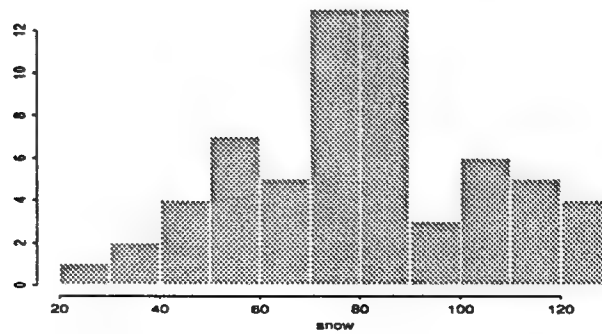
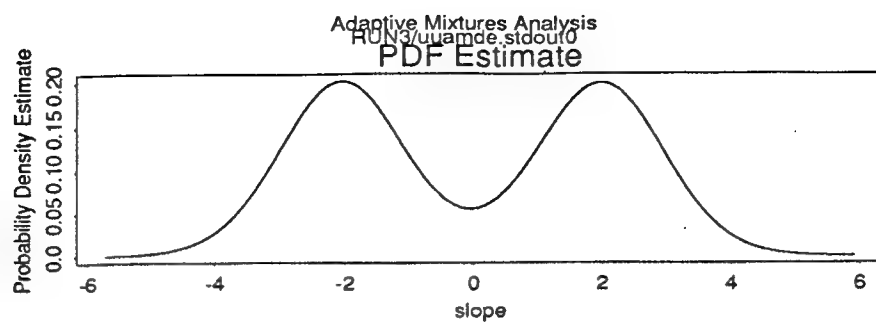
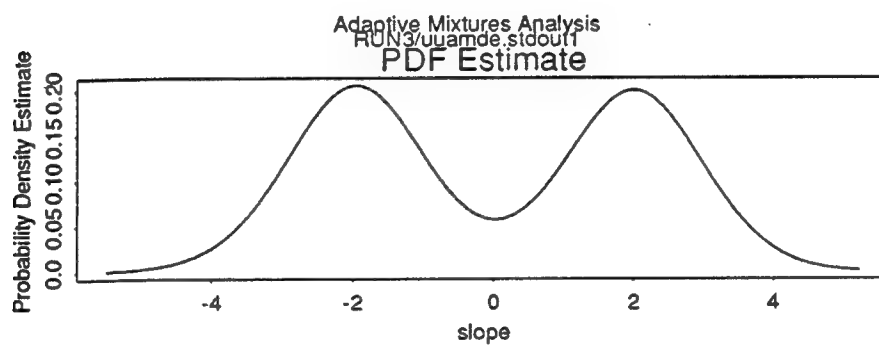
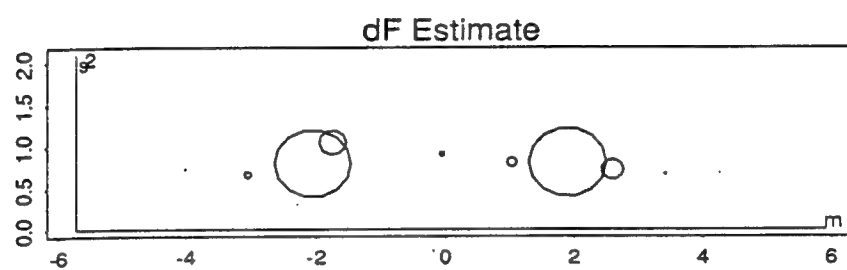


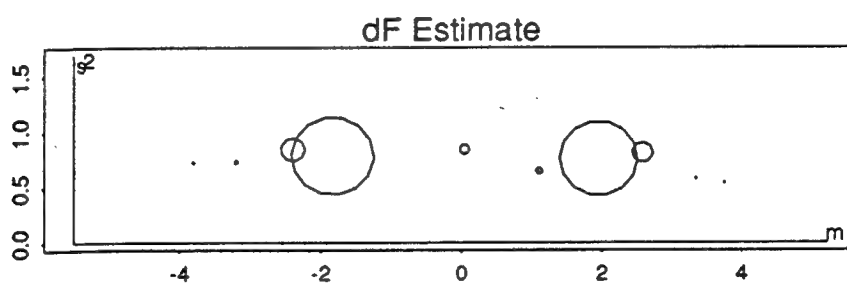
Figure 2



(a)



(b)



Figures 3 a and b.

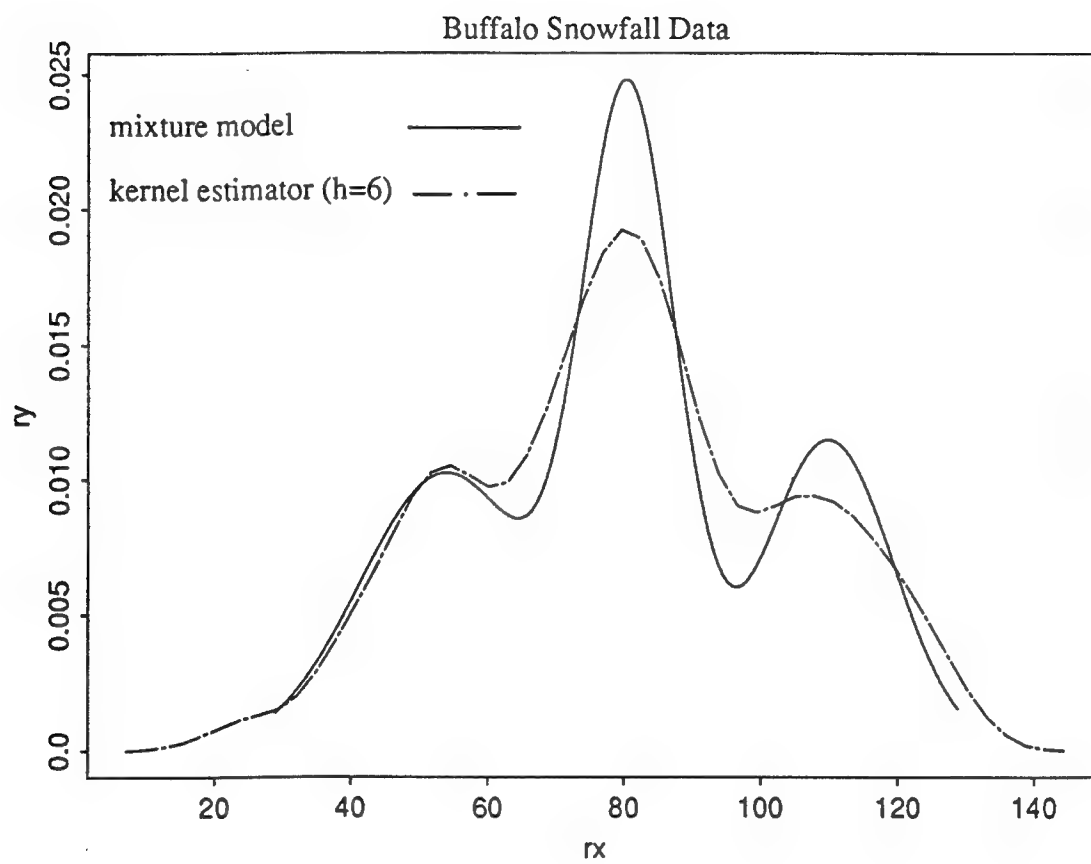


Figure 4a

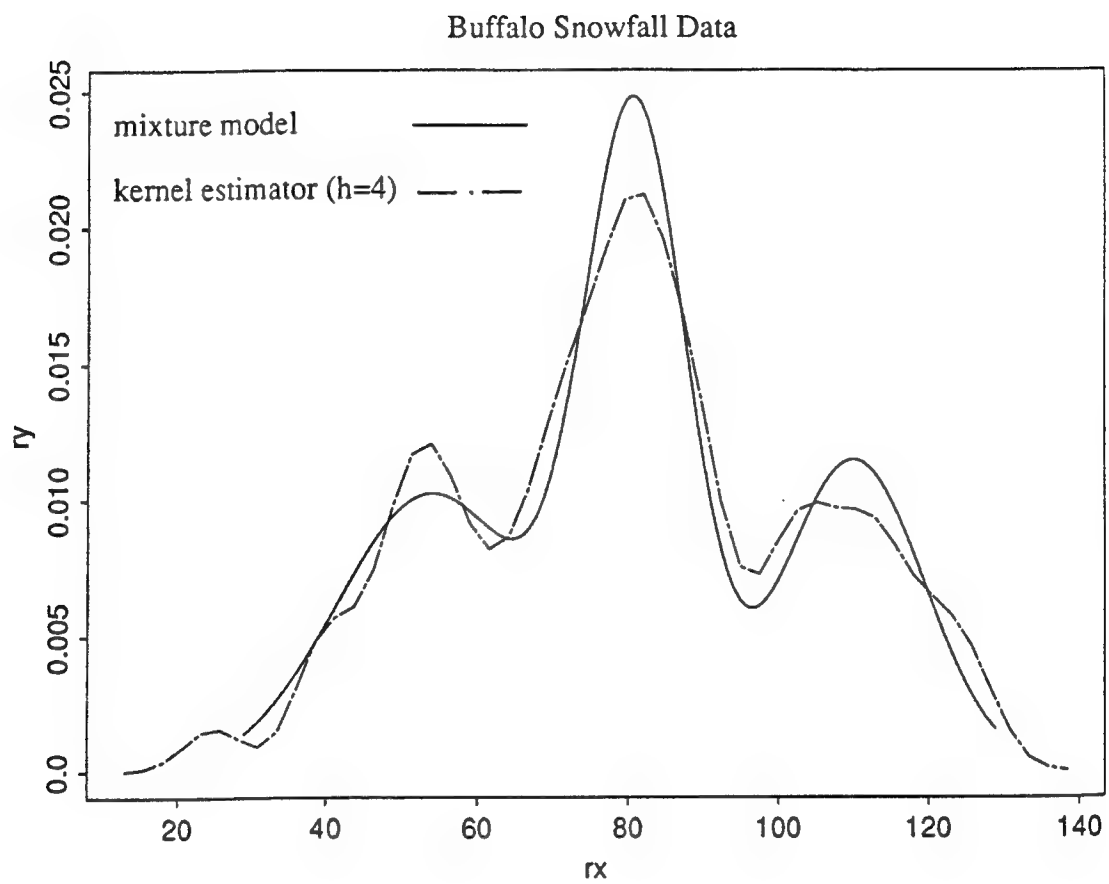
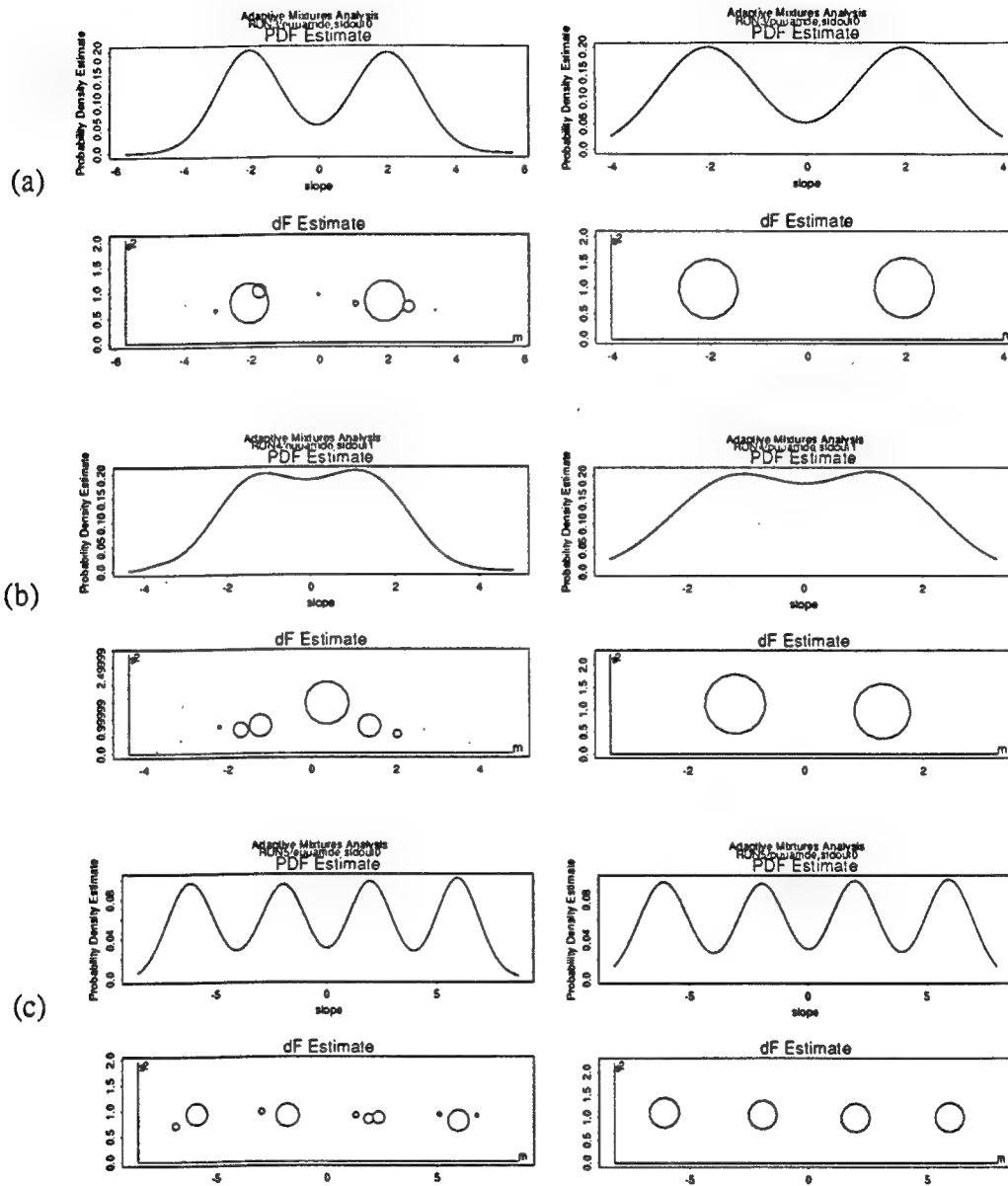


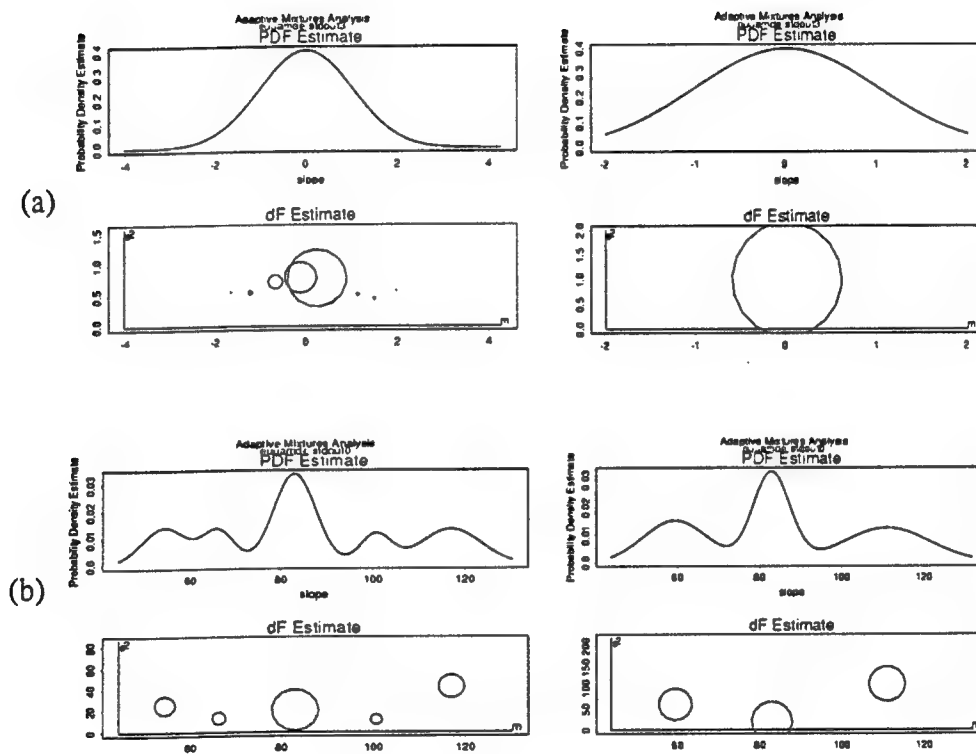
Figure 4b

EMed AMDE

PRUNED



Figures 5 a, b, and c.



Figures 6 a and b.

Table 1:

Case/# Terms	1	2	3	4	5	6	7
a - Separated Bimodal		11	4	6	2	2	
b - Bimodal		7	3	6	3	5	1
c - Quadmodal				17	5	3	
d - Standard Normal	3		9	4	4	5	
e - Buffalo Snowfall			4	10	7	4	

Table 2:

Case/# Terms	1	2	3	4	5	6	7
a - Separated Bimodal		.017	.025	.023	.025	.043	
b - Bimodal		.021	.028	.033	.030	.032	.017
c - Quadmodal				.042	.046	.063	
d - Standard Normal	.011		.024	.024	.020	.018	
e - Buffalo Snowfall							

FIGURE AND TABLE CAPTIONS

Figure 1 - Pruning curves for the reduction of a 10 term model to a 2 term model.

Figure 2 Test Cases

(a) - $\alpha(x) = .5*N(-2,1) + .5*N(2,1)$,

(b) - $\alpha(x) = .5*N(-1.25,1) + .5*N(1.25,1)$,

(c) - $\alpha(x) = .25*N(-6,1) + .25*N(-2,1) + .25*N(2,1) + .25*N(6,1)$,

(d) - $\alpha(x) = N(0,1)$,

(e) - The Buffalo Snowfall Data

Figures 3 a and b - Adaptive mixtures estimates for two of the resamplings of the data set drawn from case a.

Figures 4 a, b - Comparison of the pruned 3 term model which has been expectation maximized against kernel estimates of the original Buffalo snowfall data with bandwidths of 6. and 4.

Figures 5 a, b, and c - Expectation maximized adaptive mixtures estimates along with the output of the pruning process for the first three cases.

Figures 6 a and b - Expectation maximized adaptive mixtures estimates along with the output of the pruning process for the last two cases.

Table 1: Number of terms for each case.

Table 2: Average L_1 error for each test case for each model complexity.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE October, 1994	3. REPORT TYPE AND DATES COVERED Technical	
4. TITLE AND SUBTITLE A Method to Determine the Structure of an Unknown Mixture Using the Akaike Information Criterion and the Bootstrap			5. FUNDING NUMBERS N00014-92-J-1303 N00014-93-1-0527	
6. AUTHOR(S) Jeffrey L. Solka, Edward J. Wegman, Carey E. Priebe Wendy L. Poston and George W. Rogers				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Center for Computational Statistics George Mason University Fairfax, VA 22030			8. PERFORMING ORGANIZATION REPORT NUMBER TR no.105	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Department of the Navy Office of the Chief of Naval Research Mathematical Sciences Division 800 N. Quincy Street Code 1111SP Arlington, VA 22217-5000			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Navy position, policy, or decision, unless so designated by other documentation.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) Given i.i.d. observations $x_1, x_2, x_3, \dots, x_n$ drawn from a mixture of normal terms one is often interested in determining the number of terms of the mixture and their defining parameters. Although the problem of determining the number of terms is intractable under the most general assumptions there is hope of elucidating the mixture structure given appropriate caveats on the underlying mixture. This paper examines a new approach to this problem based on the use of Akaike Information Criterion (AIC) based pruning of data driven mixture models which are obtained from resampled data sets. Results of the application of this procedure to artificially generated and real world data sets are provided.				
14. SUBJECT TERMS AIC, pruning, mixture models, resampling			15. NUMBER OF PAGES 28	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
October, 1994

3. REPORT TYPE AND DATES COVERED
Technical

4. TITLE AND SUBTITLE

A Method to Determine the Structure of an Unknown Mixture Using the Akaike Information Criterion and the Bookstrap

5. FUNDING NUMBERS

DAA L03-91-G-0039

DAAH04-94-G-0267

6. AUTHOR(S)

Jeffrey L. Solka, Edward J. Wegman, Carey E. Priebe
Wendy L. Poston and George W. Rogers

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center for Computational Statistics
George Mason University
Fairfax, VA 22030

8. PERFORMING ORGANIZATION
REPORT NUMBER

TR no.105

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING
AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Given i.i.d. observations $x_1, x_2, x_3, \dots, x_n$ drawn from a mixture of normal terms one is often interested in determining the number of terms of the mixture and their defining parameters. Although the problem of determining the number of terms is intractable under the most general assumptions there is hope of elucidating the mixture structure given appropriate caveats on the underlying mixture. This paper examines a new approach to this problem based on the use of Akaike Information Criterion (AIC) based pruning of data driven mixture models which are obtained from resampled data sets. Results of the application of this procedure to artificially generated and real world data sets are provided.

14. SUBJECT TERMS

AIC, pruning, mixture models, resampling

15. NUMBER OF PAGES
28

16. PRICE CODE

17. SECURITY CLASSIFICATION
OF REPORT

UNCLASSIFIED

18. SECURITY CLASSIFICATION
OF THIS PAGE

UNCLASSIFIED

19. SECURITY CLASSIFICATION
OF ABSTRACT

UNCLASSIFIED

20. LIMITATION OF ABSTRACT

UL